The feasibility of testing whether *Fasciola hepatica* is associated with increased risk of verocytotoxin producing *Escherichia coli* O157 from an existing study protocol

GL Hickey, PJ Diggle, TN McNeilly, SC Tongue, ME Chase-Topping, DJL Williams

INTRODUCTION

The parasite *Fasciola hepatica* is a major cause of economic loss to the agricultural community worldwide as a result of morbidity and mortality in livestock, including cattle. Cattle are the principle reservoir of verocytotoxigenic *Escherichia coli* O157 (VTEC O157), an important cause of disease in humans. To date there has been little empirical research on the interaction between *F. hepatica* and VTEC O157. It is hypothesised that *F. hepatica*, which is known to suppress type 1 immune responses and induce an anti-inflammatory or regulatory immune environment in the host, may promote colonisation of the bovine intestine with VTEC O157. Here we assess whether it is statistically feasible to augment a prospective study to quantify the prevalence of VTEC O157 in cattle in Great Britain with a pilot study to test this hypothesis.

METHODS

On observing the data, we will fit a mixed effects logistic regression model. In the absence of data on other explanatory variables, this model will be

$$\text{logit}(p_{ij}) = \alpha_j + \beta x_{ij}$$

where

- p_{ij} is the probability of cow i on farm j testing positive for VTEC O157
- α_j is the intercept for farm j, such that each α_j are conditionally independently distributed normally with mean μ and standard deviation σ
- x_{ij} = 1 if cow i on farm j tests positive for *F. hepatica*, and 0 otherwise
- β is the natural log odds ratio (OR) for a positive *F. hepatica* test

There is no closed-form solution for the power of the test, and we therefore use a simulation-based approach (Gelman and Hill 2007) as follows.

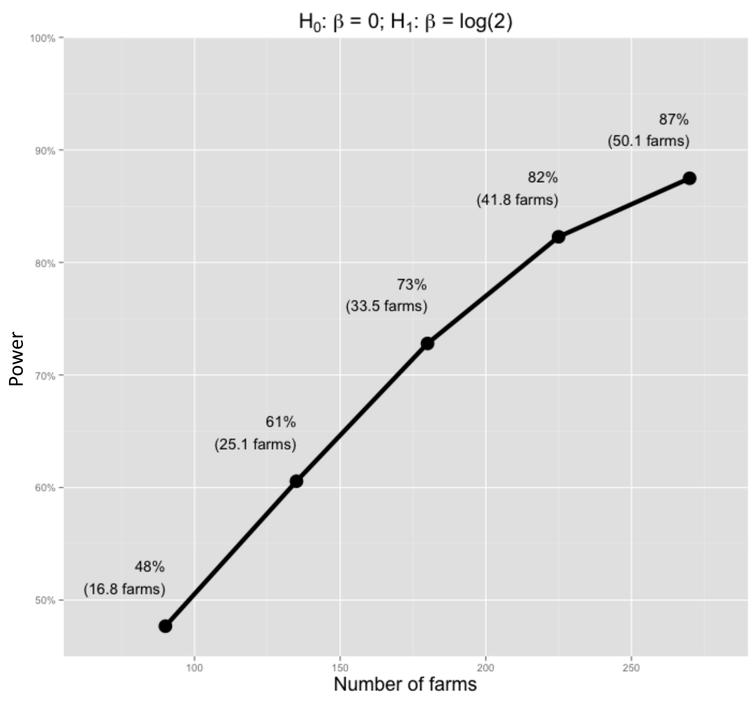
1. Simulate a plausible synthetic dataset that adheres to any known constraints under the alternative hypothesis.
2. Fit the proposed regression model.
3. Test the null hypothesis at the 5% significance level.
4. Repeat 2500 times and calculate the power as the proportion of simulations where the null hypothesis was rejected.

We simulate synthetic datasets (item 1 above) using marginal prevalence data from the published literature to calculate distributional parameters by exploitation of the total law of expectations.

RESULTS

The power curve is shown below. It shows that from a potential 270 farms included in the FSA survey, only 50 farms on average, equating to an average of 1645 pat samples, would have a sample VTEC O157 prevalence of >0% or <100% and thus require testing for *F. hepatica*. This would yield power of 87% to detect an odds ratio of 2, hence there is potential to test fewer farms. Repeating the exercise with 225 farms, we find that we expect to apply *F. hepatica* testing to 42 farms, equating to approximately 269 fewer pat sample tests, whilst yielding power of 82%.

A sensitivity analysis on the power to detect different effect sizes ranged from 13.5% (for OR = 1.2) to 99.5% (for OR = 3.0). The power to detect an OR of 1.8 would be 75%. Sensitivity analyses on changes to the farm-level prevalence of VTEC O157 showed that a reduction was commensurate with a reduction in statistical power, whereas change to the *F. hepatica* prevalence values had relatively smaller influence.



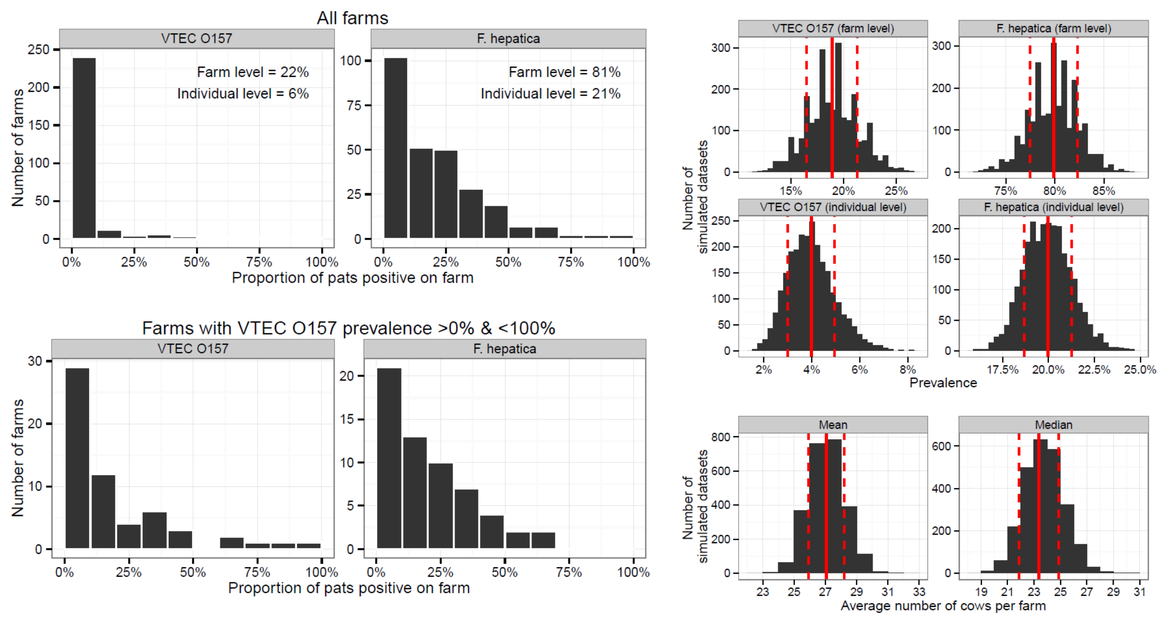
DATA SIMULATION

Farms. The proposed study design should have a mean and median sample size, n_j, of approximately 27 and 23 cows per farm respectively, with a range of 1 to 113. We simulate sample sizes $n_j - 1$ from a Beta-Binomial(112, 1.32, 4.35) distribution.

F. hepatica infection. Based on existing data, we want the approximate marginal prevalence of *F. hepatica* among individual cows (ignoring clustering effects) to be 20%, and the farm-level prevalence (the proportion of farms with ≥1 cow testing positive for *F. hepatica*) to be 80% (McCann et al. 2010). To achieve this, within each farm we infect cows with *F. hepatica*, *in silico*, with a within-farm probability r_j sampled from a Beta distribution with shape parameters 0.99 and 3.97.

Farm effects and VTEC O157 infection. Based on existing data, we want the approximate marginal prevalence of VTEC O157 among individual cows to be 4%, and the farm-level prevalence to be 19% (Pearce et al. 2009). When β = log(2), we simulate farm-level random effects α_j on the logit scale from a normal distribution with mean μ = –7.09 and standard deviation σ = 3.52. We then infect each cow i on farm j, *in silico*, with probability p_{ij}, as defined by the model above.

We exploit the fact that each VTEC O157 test result will be known in advance of the samples being requested for *F. hepatica* testing. If each sample were to be tested at the same time for both pathogens, then >7000 *F. hepatica* tests would need to be carried out across 270 farms. As VTEC O157 has a relatively low prevalence, many farms will have a sample prevalence of 0%. These farms contribute little to the estimation of the model parameters of interest; therefore we exclude them prior to fitting the regression model. By similar reasoning, we excluded the few farms with 100% VTEC O157 prevalence.



Left panel: a single simulation showing distribution of farm-level prevalence data. Right panel: distribution of average prevalence and sample size across 2500 simulations.

CONCLUSIONS

From a total of 270 farms (mean 27 cows per farm) that will be tested for VTEC O157, power of 87% can be achieved, whereby testing of *F. hepatica* would only be necessary for an expected 50 farms, thus considerably reducing costs. Pre-study power calculations are an important part of any study design. The framework developed here might be applicable to the study of other co-infections.

Our study makes a number of assumptions, including infection prevalence values holding constant; no spatiotemporal variation; and perfect diagnostic tests. Trimming the data by excluding farms with either 0% or 100% sample prevalence for VTEC O157 confers a substantial reduction in cost but could, in principle, invalidate the subsequent analysis. We recommend that users of the approach verify the properties of the study by simulation using scenarios tailored to each application.

REFERENCES & ACKNOWLEDGEMENTS

- Gelman A, Hill J, 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. *Cambridge University Press*, NY.
- McCann CM et al., 2010. The development of linear regression models using environmental variables to explain the spatial distribution of *Fasciola hepatica* infection in dairy herds in England and Wales. *Int. J. Parasitol*. 40, 1021–8.
- Pearce MC, et al., 2009. Temporal and spatial patterns of bovine *Escherichia coli* O157 prevalence and comparison of temporal changes in the patterns of phage types associated with bovine shedding and human *E. coli* O157 cases in Scotland between 1998-2000 and 2002-2004. *BMC Microbiol*. 9, 276.
- Hickey GL, et al. The feasibility of testing whether *Fasciola hepatica* is associated with increased risk of verocytotoxin producing *Escherichia coli* O157 from an existing study protocol. *Prev Vet Med*. 119:97-104. Corrigendum, in press.

VTEC O157 data used in this study was collected as part of the International Partnership Research Award in Veterinary Epidemiology (IPRAVE), Epidemiology and Evolution of Enterobacteriaceae Infections in Humans and Domestic Animals funded by the Wellcome Trust.FS101055. *E. coli* O157 super-shedding in cattle and mitigation of human risk is funded by the U.K. FSA: http://www.food.gov.uk/science/research/foodborneillness/ecoliresearch/fs101055/.

CONTACT
Professor Diana Williams
University of Liverpool
+44 (0)151 795 0220
williadj@liverpool.ac.uk